Computer Standards & Interfaces 62 (2019) 53–63

Contents lists available at ScienceDirect

Computer Standards & Interfaces

journal homepage: www.elsevier.com/locate/csi

Performance and usability tradeoff in a cluster display wall☆

Ismael Arroyo[a], Francesc Giné[⁎,b], Concepció Roig[b], Antoni Granollers[b]

[a] Davantis Technologies S.L., Autonomous University of Barcelona, Eureka Building, Bellaterra 08193, Spain
[b] Polytechnic Institute of Research and Innovation in Sustainability (INSPIRES), Universitat de Lleida, C/Jaume II, 69, Lleida 25001, Spain

ARTICLE INFO

Keywords:
Cluster display wall
Modeling
Performance
Usability

ABSTRACT

Cluster-based display walls provide cost-effective and scalable display infrastructures with high resolution and large display area, making them suitable for a wide range of high-resolution applications. As a consequence, a wide offer of new cluster display-wall platforms together with their software frameworks have been proposed. Their performance and the satisfaction of their users have aroused the interest of some researchers. This work is focused on the Liquid Galaxy cluster display wall originally built to run Google Earth to create an immersive experience for the users. In this paper, the Liquid Galaxy is benchmarked running Google Earth, as a representative interactive application with high performance requirements, in different configurations and environments, to test the satisfaction, effectiveness and efficiency. Thus, we wish to know how users react to the system performance. In order to do so, we use a performance metric defined in previous research to relate the performance of the system with the user's perception. Taking into account the trend of this metric in the experimentation, we model the behavior of the system in a way that the performance for any given visualization cluster running Google Earth could be predicted by using a reference system.

1. Introduction

A multi-display visualization environment is a system where there is more than one display or projector to increase the visualization area. The multi-displays are simply the way the information from the application is delivered to be visualized, but these can be managed differently depending on the back-end infrastructure. Consequently, they can be classified into two groups: *Display Walls and Cluster-based Display Walls*. A display wall is an infrastructure in which various screens are distributed in tiles connected to a single powerful computer integrated with multiple video outputs. Its main problem is that the images are stretched, as the resolution of a single screen is resized to fit all the screens together. On the other hand, a cluster-based display wall consists of a number of synchronized PCs where each node of the cluster has one or more displays connected to it, improving the quality of the visualization by increasing the pixel density efficiently. Thus, cluster-based displays are performance-, memory- and display-scalable, easy to maintain and upgrade. Some examples are CAVE [1], GeoWall [2], Garuda [3] or Liquid Galaxy [4].

The majority of applications executed in a cluster-based display follow two approaches: *master-slave* or *client-server*. In the master-slave applications, the data-set is mirrored across all the nodes and with multiple instances of a program running in parallel, one on each node. An example of a master-slave application is Google Earth [5]. In the client-server approach, the server runs a different instance of the program executed by each client, distributing appropriate data to each client node and performing the synchronization among the client nodes. One of many examples of such an application is CaveSL, a modified version of the game Second Life [6].

This rising trend for cluster display walls drew our attention and brought up some questions about how different kinds of applications perform with a system built up with commodity hardware, how the end-user would feel when using and experiencing these new visualization systems and how we could model both the user experience and performance acknowledging the system parameters.

In order to answer these questions, our research focused on the use of a specific cluster display wall hemisphere infrastructure developed by Google, named Liquid Galaxy (LG) [4], built up with commodity hardware. Liquid Galaxy system is a cluster infrastructure that provides an immersive visualization made up of eight displays by default, each connected to a computer node. Although the original project had eight nodes and displays, the cluster can be extended depending on the infrastructure. As a cluster, it is easily scalable as it is expandable, while minimizing cost by requiring low-cost infrastructure. Although it was

☆ This work was supported by the MEyC under contracts TIN2014-53234-C2-2-R and TIN2017-84553-C2-2-R.

⁎ Corresponding author.
E-mail addresses: ismael.arroyo@davantis.com (I. Arroyo), sisco@diei.udl.cat (F. Giné), roig@diei.udl.cat (C. Roig), antoni.granollers@udl.cat (A. Granollers).

https://doi.org/10.1016/j.csi.2018.08.002
Received 15 January 2018; Received in revised form 27 June 2018; Accepted 15 August 2018
Available online 23 August 2018
0920-5489/

Fig. 1. Example of Liquid Galaxy.

originally conceived to run the Google Earth application, other applications can be run on Liquid Galaxy, including Quake III Arena [7], WebGL applications [8] or video streaming. Fig. 1 shows the LG installed in the Technological Park in the city of Lleida (Spain) running Google Earth.

In previous research [9], we analyzed how an LG system, assembled from commodity hardware and running a set of representative applications, provided good enough performance metrics for different configurations. However, the user's point of view was missed. Thus, given that the majority of applications running in a cluster-based display wall are interactive, we are interested in knowing how users perceive this performance. Note that to the best of our knowledge, no previous work focused on cluster display walls has faced the challenge of analyzing and modeling the relationship between performance and the satisfaction of the users.

To this end, the present paper is focused on knowing how the users react to an LG system made up of different configurations running interactive applications in distinct environments. Likewise, we are interested in obtaining knowledge of user satisfaction by relating it to the performance metrics. In order to achieve this, the LG system was tested with users with different profiles through the usability attributes of satisfaction, effectiveness and efficiency [10] in different scenarios. In order to obtain the satisfaction, some tests were carried out in which users were required to respond to some post-task questionnaires about their feelings while using the system. Likewise, the system performance was monitored throughout the tests by means of a performance metric called *Visualization Rate (VR)*, designed to supply information about the efficiency and effectiveness of the system. Finally, we developed a model to approximate the performance of the platform running Google Earth as a representative example of interactive application. Our results reveal that with knowledge of the system performance, which can be found with objective metrics, the suitability of a cluster display-wall for use under certain user requirements can be estimated. This is a very encouraging result, as it can facilitate the spread of the use of the LG platform to a broad range of users and fields (education, professional, research, etc.).

This paper is structured as follows. Section 2 presents the background of cluster visualization systems. Section 3 evaluates the performance of an LG running a set of different interactive applications. Section 4 describes the Visualization Rate metric used to relate the performance with the user's perception. In Section 5, we evaluate the usability of the LG system in different contexts and for a wide set of user profiles. Section 6 defines a theoretical schema used to model the behavior of the LG system running Google Earth. Finally, Section 7 concludes the paper and discusses future directions.

2. State of the art

As our aim is to extend the use of LG to a wide range of environments by providing a low-cost cluster display-wall infrastructure, we wanted to focus on the side of the user will be using such environments.

For this reason, we focused on works in the literature that analyze the usability of the different cluster display walls by using the most recent usability standard [11] that is defined as *"the degree to which a product or system can be used by specific users to achieve specified goals with effectiveness, efficiency and satisfaction in a specific context of use. Usability can either be specified or measured as a product quality characteristic in terms of its sub-characteristics, or specified or measured directly by measures that are a subset of quality in use"*.

Some of these works are focused on analyzing the satisfaction of users of a cluster display wall compared with desktop systems.

Tan et al. [12] studied how their system, called Infocockpit, can make information memorable. Infocockpit is a projector-based display wall with ambient visual and auditory displays that engage human memory for location. In their work, the authors made users complete semantic tasks that consisted of remembering pairs of words and then recalling them on both Infocockpit and a single desktop computer. They concluded that Infocockpit improved the memory of the user in comparison with a desktop computer.

Ball and North [13] studied the satisfaction from a 3x3 large tiled display compared with two smaller displays. They tested both environments with a task with quantitative results based on finding targets of different sizes together with observations that the users made during the test. They concluded that display walls, which users have to walk along in order to visualize the data, significantly outperform smaller displays that use pan and zoom navigation.

Likewise, some works from the literature focus on analyzing the performance of cluster display walls in terms their of efficiency and effectiveness.

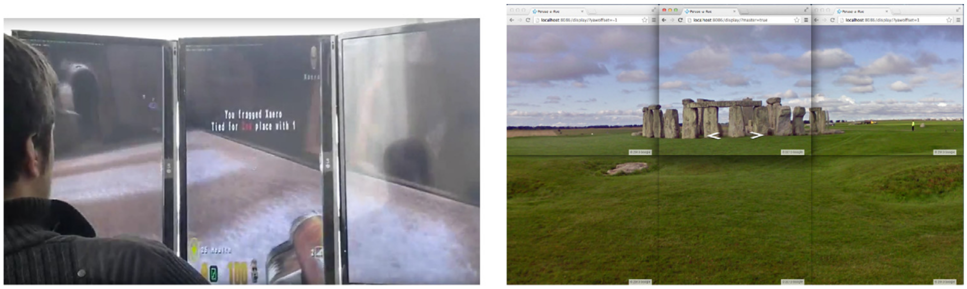
Humphreys et al. [14] describe their proposal to render OpenGL applications using their middleware WireGL and compare its performance with another middleware named Broadcast [15]. They ran three different benchmarking applications over a variable distribution of displays ranging from 1×1 to 8×4 devices, while monitoring the performance in frames per second. They stated that their system was somewhat slower when the distribution was 1×1 but it maintained the performance across all configurations whilst the Broadcast middleware did not.

Neal et al. [16] studied the performance of ClusterGL in comparison with Chromium and BroadcastGL in a network bandwidth constrained environment. For their work, they used different optimization techniques and noted the impact of each individual technique and also the combination of them. They were able to determine the frames per second increase of ClusterGL against the other middlewares and the performance upgrade from the optimization techniques.

Despite the growing literature devoted to the study of the performance and usability of cluster display wall systems [17–21], to the best of our knowledge, there are no studies that focus on establishing and modeling the relationship between the performance of the system and the satisfaction of its users. Therefore, our main goal is to relate some usability aspects of the User Experience [22] with the performance of the LG cluster.

3. Liquid Galaxy performance based on the interactive application model

This section analyzes the main performance metrics of CPU, Memory and internal/external network traffic of an LG system assembled from commodity hardware running a representative set of Client-Server (C-S) and Master-Slave (M-S) interactive applications. It is worth pointing out that experimentation is only focused on interactive applications given the aim of analyzing the relationship between performance and usability of a cluster-based display wall. A homogeneous LG made up of 3 nodes was used, where each node is composed of an Intel Core i53GHz, with 2x4GB RAM 1600 MHz, SSD 128GB, NVIDIA GT620 and a32" screen. All the tests were executed in a closed laboratory and the computing metrics were monitored by means of the



(a) Quake III Arena (b) Peruse-a-Rue

Fig. 2. Examples of Client-Server applications.

following tools: *Top*, which was used to monitor the CPU and RAM, and *Tshark*, a command-line based Wireshark version, which is a packet sniffer used to monitor the network traffic.

In the case of the client-server approach, we executed the two following applications:

- Videogame Quake III Arena (Fig. 2a). To provide an immersive experience, all the nodes were connected to a local server. One node in the middle was the actual player, while the other nodes were just spectators of this player and showed a view offset depending on their physical position inside the cluster. The test lasted 120 s and was performed in the Q3DM6 map with 5 players.
- Peruse-a-Rue (Fig. 2b). This application allows the surroundings of a specific place to be visualized using 360 ° photos, when navigating along the street. The test developed with Peruse-a-Rue consisted of a tour through a street in New York city. From a specific point, the application moved the visualization point forward every 5 s, until it stopped 60 s later.

The testing of the master-slave approach was done by monitoring the two following interactive applications:

- Google Earth. The well-known application developed by Google [5] that allows visualization of places all over the globe. The test was performed for a tour of 8 points of interest in the city of Barcelona (Spain), where the environment to be visualized contains a high-density of 3D buildings and so, high requirements for computing resources. The time interval between consecutive jumps was set at 30 s, as it permitted whether the imagery had been completely loaded to be analyzed.
- Point Cloud Viewer. This is a web-based application developed ad-hoc for the LG to visualize different types of cloud point files [23]. The navigation through a specific point cloud was carried out using WebGL on Chromium browsers. A navigation through a figure composed of a point cloud (a set made up of an enormous quantity of points to represent something), which was captured by a local camera, was monitored over 60 s.

For all tests performed with these interactive applications, Table 1 shows the Arithmetic Mean ($\overline{x}$) and Standard Deviation (σ) for each

Table 1
Performance evaluation.

Applications		%CPU (MB)		Mem. (KB)		Ext.Net. (KB)		Int.Net.	
Model	Name	$\overline{x}$	σ	$\overline{x}$	σ	$\overline{x}$	σ	$\overline{x}$	σ
C-S	Quake III Arena	22	1	110	10	0	0	30	10
C-S	Peruse-a-Rue	2	0.2	50	5	200	44	60	16
M-S	Google Earth	41	22	1390	685	534	245	687	350
M-S	Cloud Viewer	25	5	456	65	0	0	320	30

monitored metric. These results reveal that C-S applications have stable behavior given the low standard deviation obtained for all the metrics. Likewise, the arithmetic mean reveals that these applications are characterized by a low consumption of computing resources. In relation to the M-S applications, the arithmetic mean reveals that the requirements for computing resources increase, especially in the case of Google Earth. However, according to the results for CPU and Memory, it can be seen that there is no need for high-performance nodes, given that the highest average CPU usage was below of 50% and Memory below 1.5GB. The network analysis shows that a key performance parameter is the network usage because Google Earth has a heavy request for data from Internet and many internal synchronization packets per second, which would increase if the system was scaled higher than 3 nodes. Thus, the external network bandwidth is the real bottleneck in the system. In addition, the high deviation achieved by each parameter for the Google Earth is surprising. This is given by the fact that the system is stressed whenever the user downloads a new image, while the rest of the time, the system resources consumption falls to near zero.

Therefore, Google Earth is the interactive application that is most sensitive to the performance issues. According to this, this application was chosen as a representative benchmark throughout this paper.

4. Relating performance to user perception

In our previous research [9], a new performance metric was defined, called Visualization Rate (VR), that gives knowledge of when the cluster display wall has loaded all the visual elements. The VR is the average CPU idle time for a cluster of n nodes and is calculated with the following equation:

$$VR = \frac{100}{n} \sum_{i=1}^{n} \frac{T_{idle_i}}{T_{total}}, \tag{1}$$

where T_{total} is the total time of the test and T_{idle_i} is the time when the CPU load of node n_i is below a minimum threshold.

Note that a CPU load below this threshold means that the CPU is idle and the images have been fully loaded. This procedure is illustrated in Fig. 3, where the blue line near 5% of CPU usage is a threshold to indicate when the application is running in an idle state. The peaks depicted in this Figure correspond to when the CPU is processing images, composed by polygons and textures, from Google Earth, and the values that drop below the threshold represent when the CPU becomes idle. The CPU usage information was gathered every second from the information given automatically by the system monitor. Note that when the VR is near 100%, this denotes a high visualization rate and so, we assume this implies a good user perception as images are fully loaded, while having a VR equal to 0% means that the data has not been fully loaded and, thus, it has been ineffective for the users' feelings of satisfaction. Any value between 0% and 100% indicates how efficiently the system has performed. It is worth pointing out that our reasoning assumes that people are navigating to visualize a specific point on the Earth.

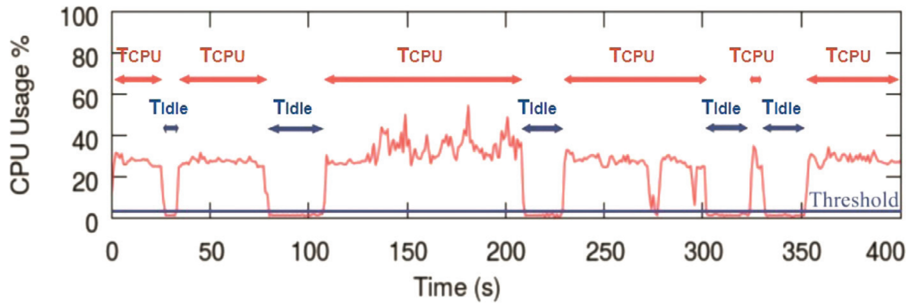


Fig. 3. Example of CPU usage.

Taking the above reasoning into account, it can be assumed that VR constitutes a metric that not only gives information about performance, but can also relate this performance to give knowledge of user perception. It is worth remarking that the decision to calculate the VR using the CPU instead of other common metrics used in the literature, such as frames per second or latency, came following a study about how the Google Earth application behaved when displaying the multimedia information. In this study, we analyzed different metrics in order to find a reliable way to ensure that the buildings and images displayed by the multimedia application were completely loaded. By looking at the common GPU metric of frames per second (fps), we could only determine that the application was running smoothly as it was displaying constantly 30 fps throughout the tests. Therefore, we concluded that this metric was not determinant for our purpose. On the other hand, the CPU load yielded information about how much time the system needed to load all the multimedia information because when the imagery was completely drawn, the CPU load dropped to idle values. Note that this multimedia information was basically the number of polygons needed to draw the scenario and buildings and also the images or textures that are used to ”paint”, or render, such polygons. For this reason, the CPU is the main direct metric that we use to calculate the VR.

After defining the VR metric, and in order to analyze what the minimum acceptable VR is, a discount usability test [24] that was reported in [9] was carried out. It was performed with 86 volunteers of different ages and Multimedia eXperience Level (MXL). They were visitors to the *Euskal Encounter fair* (https://www.euskal.org/). The test consisted of navigating to six different points of interest in New York automatically, in a fully guided tour. The user had to press a key when he/she wanted to go on to the next point on the tour. The results obtained in this experiment led us to conclude that in conditions of fully directed interactivity, a VR value of 30% could be assumed as the minimum acceptable to provide users with navigating and visualizing experience.

5. Usability analysis of the Liquid Galaxy platform

In line with the aims of the present work, we performed some experiments in which participants navigated throughout different routes in Google Earth with different degrees of autonomy, so that we were able to analyze the relationship of users' interactivity with the system performance. So, the LG system was tested with users for the usability attributes of satisfaction, effectiveness and efficiency in two different scenarios. In our context, we understand that satisfaction is a subjective parameter that measures the users' perception, effectiveness is the ability of the system to load all the images while running the application and efficiency is led by the time required to load these images.

Two tests with different levels of freedom to achieve the proposed objective were carried out.

- A *Driven test*, where users navigated autonomously through a series of guided and predefined tasks supplied by a facilitator.
- A *Field test*, that was carried out in a travel agency where clients (users) were encouraged to navigate freely to wherever they would like to visit.

In both tests, users had to answer some post-task questions about how they felt after completing each task. At the same time, the VR system performance metric was monitored throughout the tests.

Beforehand, users were asked about their Multimedia eXperience Level (MXL) by choosing a value between 1 and 5 (1 being the lowest value and 5 the highest) that indicates the familiarity of the users with the use of multimedia applications.

In all tests, a facilitator guided the participants. Once the test started, the facilitator could only answer specific questions or give subtle advice when the participant was struggling to complete a task for a long time. All tests were done with homogeneous LG systems made up of three up to eight nodes (i5 3330, NVIDIA GeForce GT620, 8GB RAM, SSD) with the Squid disc cache enabled. The interaction device was the 3D Space Navigator (circled in Fig. 1). As we knew that this is different and more complex than a common mouse, we dedicated some minutes to instructing the participants about how to use the device before starting the test.

5.1. Driven test

With the LG infrastructure, we carried out two tests that consisted of navigating to some well-known places around the world with an LG with two different node composition.

The environment in which both tests were performed was a room equipped with an LG system connected to a gigabit local network with a broadband connection of 8Mbps. The first test was done with an LG system with 3 nodes and 27 volunteers, 15 females and 12 males, which ages ranged from 12 to 68. In the second test, the LG system was composed of 8 nodes and there were 25 volunteers participants, 13 males and 12 females with ages ranging from 11 to 65, with a similar profile to those in the first test. All the volunteers were informed about the system, their role in the test and their privacy rights according to Spanish law (everyone signed a consent form in order to keep them). Table 2 shows the MXL of the people involved in both tests, grouped by age ranges.

In order to acquire specific data related to user behavior, an eye tracking device was used. This technology adds a complementary view to researchers as it allows us to understand what users look at and what

Table 2
MXL in relation to age ranges.

	MXL (3 Nodes)					MXL (8 Nodes)				
Age Range	1	2	3	4	5	1	2	3	4	5
11–16	1	–	1	2	–	–	–	1	2	–
17–21	–	–	–	1	2	–	–	–	1	3
22–35	3	2	–	3	1	2	1	1	3	2
36–68	5	2	2	2	–	4	3	1	1	–

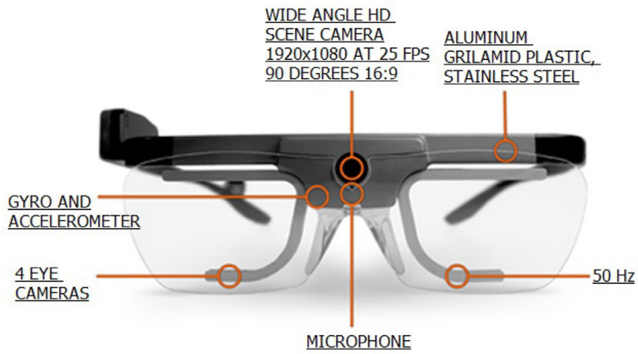


Fig. 4. Tobii ProGlasses 2.

they do not see when interacting with a display interface. The specific device that was used is the Tobii ProGlasses 2 [25], shown in Fig. 4 that allows us to record the users' eye activity across multiple screens.

While wearing the Tobii Glasses, each participant was asked to do the following four tasks, namely navigating to one of the following well-known places in the world:

- T1: New York (USA): Statue of Liberty.
- T2: Barcelona (Spain): Football Club Stadium.
- T3: Sydney (Australia): Bay of the Opera House.
- T4: Lleida (Spain): City where the participants live.

T1, T2 and T3 were made up of two sub-tasks. In the first sub-task (Tia), the system positioned itself automatically at the first place in the city, so that participants had to answer a simple related question. This way, the participants did not use the controller for the first part of each task. However, in the second sub-task (Tib), they had to use the controller to reposition the view and move around to be able to answer the related question. Task T4 consisted of a free flight from Sydney (T3b) to a specific point of interest in Lleida. Fig. 5 shows the pictures of the 7 sub-tasks making up the test.

Table 3 shows the questions that users had to answer to complete each sub-task. Although the answers were not so important in themselves, this was a way of forcing the user to interact with the system and show interest when doing the tasks. With the aim of acquiring information to evaluate the UX with the minimum set of questions, we additionally issued the two following questions to the participants after each task:

- Q1: "Of the following drawings, mark which one best explains how you felt when performing the task". The drawings, based on the *LemTool* emotional scale, consisted of the eight icons in Fig. 6 that represent four positive and four negative emotions.
- Q2: "From 0 to 10, mark your satisfaction with the load timeout of the images". The possible answers were "0, 2, 4, 6, 8 or10". The aim of this question was to obtain a fast and first-hand opinion related to the satisfaction with the image loading time, which is directly related to the VR parameter.

By collecting the data from the eye tracker and processing the answers to the questions given to the participants, we obtained a set of qualitative measures of the test. The heat maps of the eye tracker showed that users always looked at a single spot and they moved the target to the central screen of the LG, whenever possible. This behavior for the specific case of Task 2.b with 3 nodes can be seen in Fig. 7.

Table 4 shows the percentage of users with each kind of emotion (question Q1) when performing the tasks in New York (NY), Barcelona (BCN), Sydney (SYD) and Lleida (LL). Furthermore, Table 5 shows the feelings of the users about the waiting time for each of these tasks (question Q2).

We can see that the majority of the participants provided positive responses when using the system while doing the tests whenever everything was functioning and the waiting times were short. In general, 90% of the participants with both node compositions gave positive responses to the tours of NY, SYD and LL, with a few cases of discomfort. However, this percentage was lower in BCN because the user had to move the Google Earth into a view showing a significant portion of the city's buildings, thus forcing the application to download a considerable amount of data. Because of this, some people felt that they had to wait much longer than in other tasks, especially the most demanding users, including the youngest participants or those with higher technological knowledge. As a consequence, the results from BCN in Table 4 show that 14.2% of participants with 3 nodes and 20% with 8 nodes had negative feelings. This correlates with question Q2 about the waiting time shown in Table 5, where, in the case of BCN, 21.3% (3 nodes) and 32% (8 nodes) considered the waiting time unacceptable (values lower than 5). Despite some people being frustrated by the wait, they answered question Q1 more positively than expected. This leads us to think that they were enthusiastic about the system as it was new and fun for them. This last conclusion is also reflected in the comparison of the Q1 results in relation to the number of nodes. The majority of tours obtain slightly better emotions with 8 nodes than 3 nodes, which means that users appreciate the immersive environment given by the 8 nodes. Note that in the BCN tour with 8 nodes, the Q2 answers are worse given that the waiting time to load all the images is increased with the number of nodes. This happens because the broadband connection is a fixed resource and the bandwidth has to be divided between the nodes.

The Visualization Rate (VR) metric was used to obtain the effectiveness and efficiency of the system, the system performance metric was monitored throughout the test, but only the tasks where comparisons could be made (T1a (NY), T2a (BCN) and T3a (SYD)) were recorded. Those tasks were the ones that were guided, because Google Earth follows the same path from one place to another in a guided task independently of the user, thus providing fully comparable values.

Table 6 shows the percentage of VR categorized into VR ranges for both node compositions obtained by different users for the first part of NY, BCN and SYD tours (Tia). For the cases of NY and SYD, the highest VR values were near 28% with 3 nodes and 18% with 8 nodes. However, in the case of BCN, the VR achieved lower values. This is due to the fact that in BCN (Task 2a) many 3D buildings had to be rendered in order to view the desired place and, as a consequence, the loading time was higher and the users did not need to have the images fully loaded to answer the question. This is especially clear in the test with 8 nodes. In general, the VR metric is very sensitive to the number of nodes, given that the overall VR decreases when the number of nodes increases. So, the acceptable number of nodes is determined by the type of Internet connection and the amount of data to be processed. Another point to highlight is that all VR values were below 30%. This is because, in our test, the user only wanted to answer the questionnaire, and he/she was not interested in the specific imagery. As the heat map of the eye-tracking device indicated, the users' view was focused on a specific point, the one that enabled the reported question at each task to be answered. Therefore, it can be assumed that in a free flight around points of real interest by a given user, he/she would spend more time looking at a specific point, which would increase the VR metric.

5.2. *Field test*

This test was done in a real environment, a travel agency, where users could visit the places they were intending to travel to before-hand. Thus, each person had their own interest in completing the objective. Then, the user tasks were real tasks, with their own goals and motivations, that made the test results more accurate. This test was very similar to the one described in Section 5.1 to enable comparison between both tests.

The test was performed by 21 participants (12 women and 9 men) in

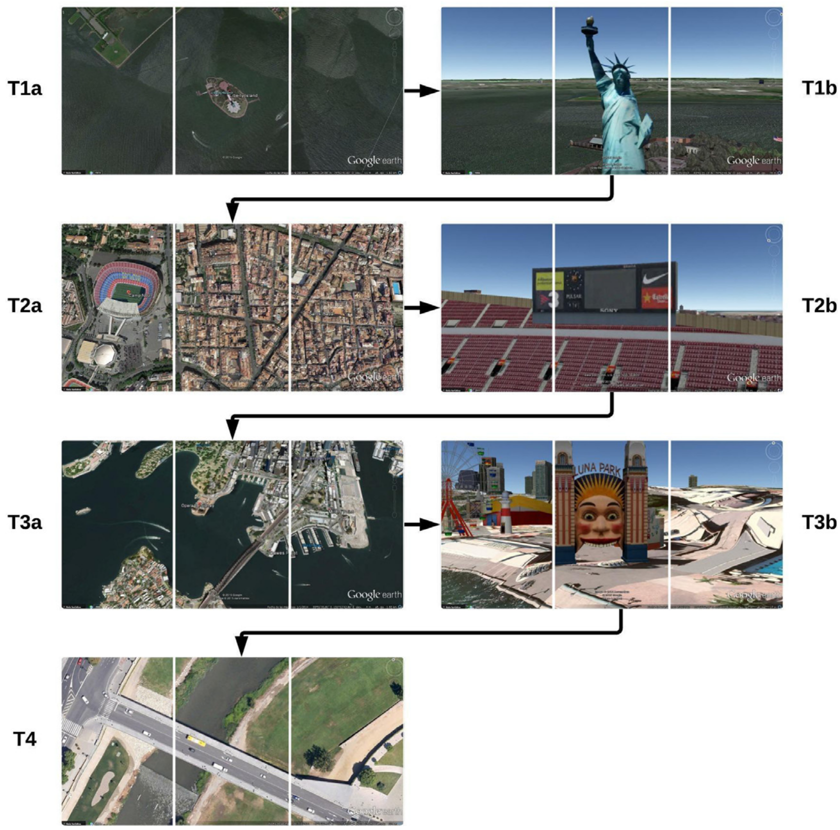


Fig. 5. Tour Flowchart - T1a Statue of Liberty from above - T1b Statue of Liberty's crown - T2a Barcelona Football Club stadium - T2b The stadium's screen - T3a Sydney Harbour - T3b Luna Park - T4a Lleida bridge. Each picture is the composition of the three screens of the system.

Table 3
Description of test tasks.

Sub-task	Position	Questions
T1a	Auto.	How many vertices does the base of the Statue of Liberty have?
T1b	Manual	How many points does the crown of the Statue of Liberty have?
T2a	Auto.	What does it say on the Barcelona Football Club stadium stands?
T2b	Manual	What make is the screen in the Barcelona Football Club stadium?
T3a	Auto.	How many buildings is the Sydney Opera House made up of?
T3b	Manual	Find a structure with a clown's face on it. What does it say?
T4	Manual	How many buses are crossing the bridge in front of the cathedral?

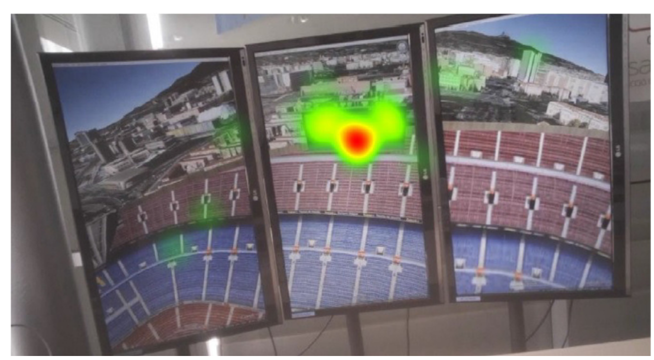


Fig. 7. Heat Map of the Task 2.b.

an LG system made up of five nodes connected to a gigabit local network with a broadband connection of 5Mbps. The multimedia skills of the users were balanced and the ages ranged from 21 to 70. Table 7 shows the MXL of the people involved in the test, grouped by age. We observed that the travel agency tends to have adult customers, which is why there were people with lower MXL values than in the previous Driven test.

The test involved a questionnaire with the same questions Q1 and Q2 described in Section 5.1. To report the results of this test, we separated the environments into high (above 160 MB), medium (between 160 MB and 110 MB) and low data density locations (below 110 MB), as every user chose a different place to visit. Therefore, the results show the average achieved in the different environments visited by the users in this test. The number of places in each category was as follows: 15

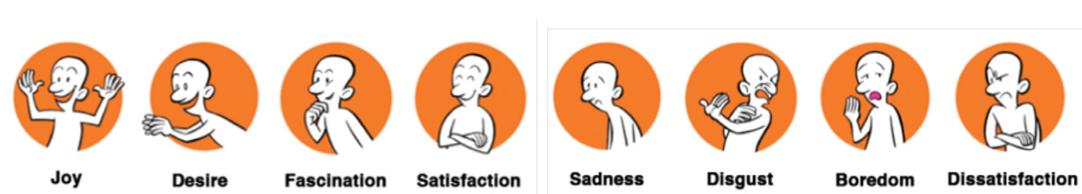


Fig. 6. Emotional choices (source LemTool).

Table 4
Results of question Q1 for the Driven test.

Tour	Nr. Nodes	Joy	Desire	Fascin.	Satisfac.	Sadness	Disgust	Bored.	Disatis.
NY	3	7.1%	7.1%	35.7%	42.9%	7.1%	0%	0%	0%
	8	24%	16%	20%	32%	4%	0%	4%	0%
BCN	3	7.1%	14.3%	21.4%	42.9%	0%	7.1%	0%	7.1%
	8	16%	12%	28%	24%	12%	8%	0%	0%
SYD	3	7.1%	0%	28.6%	57.1%	0%	0%	0%	7.1%
	8	24%	20%	16%	36%	0%	0%	4%	0%
LL	3	7.1%	14.3%	35.7%	42.9%	0%	0%	0%	0%
	8	20%	40%	18%	12%	0%	0%	0%	0%

Table 5
Results of question Q2 for the Driven test.

Tour	Nr. Nodes	0	2	4	6	8	10
NY	3	0%	0%	0%	14.3%	42.9%	42.9%
	8	0%	0%	0%	12%	52%	36%
BCN	3	7.1%	7.1%	7.1%	28.6%	42.9%	7.1%
	8	0%	12%	20%	52%	16%	0%
SYD	3	0%	7.1%	7.1%	7.1%	50%	28.6%
	8	0%	0%	4%	36%	40%	20%
LL	3	0%	0%	7.1%	14.3%	35.7%	42.9%
	8	0%	0%	0%	12%	28%	60%

Table 6
VR Ranges for T1a, T2a and T3a sub-tasks.

	VR Ranges				
Tour	Nr.Nodes	0–5%	5–10%	10–15%	15–30%
NY	3	28.6%	42.9%	21.4%	7.1%
	8	32%	56%	8%	4%
BCN	3	57.1%	35.7%	7.1%	0%
	8	68%	28%	4%	0%
SYD	3	14.3%	57.1%	21.4%	7.1%
	8	18%	52%	26%	4%

Table 7
MXL in relation to age ranges.

	MXL				
Age Range	1	2	3	4	5
17–21	–	2	–	–	–
22–35	–	2	3	–	–
36–68	4	3	2	4	1

high-density locations, 12 medium-density locations and 15 low-density locations.

Table 8 shows the kind of emotion (question Q1) when performing the tests in the different environments chosen by the users. These results show that all the participants had positive feelings when performing the tasks. Likewise, Table 9 shows the feelings of the users about the waiting time for each task (question Q2). As can be observed, the users had positive responses about the waiting times of the system. It is remarkable and expected that for the low data density locations, there was a higher value of satisfaction than for the other two types of environment.

Table 9
Results of question Q2 for the Field test.

Data Density	0	2	4	6	8	10
High	0%	0%	0%	0%	66.7%	33.3%
Medium	0%	0%	0%	0%	62.5%	37.5%
Low	0%	0%	0%	11.1%	22.2%	66.6%

We again find that users were enthusiastic about the system and they also found it more satisfying as they could visit places that truly interested them instead of locations given by the facilitator. This is reflected in the values of both tables as higher satisfaction was achieved than in the previous test.

Table 10 shows the VR values, categorized into ranges, obtained by the different types of environments. As can be seen, higher density environments tended to have lower VR values, which was expected. It is worth pointing out that all the VR values were again below 30%. In comparison with Table 6, the values obtained in this test were lower, which was due to the bad broadband connection available in the travel agency.

5.3. Discussion

The results obtained in the Driven and Field tests suggest that there might be a relation between the VR metric and questions Q1 and Q2. In order to corroborate this, the possible relation between the performance (VR) and satisfaction (Q1 and Q2) parameters was studied from a statistical point of view.

Table 11 shows the linear correlations (r) for the answers to Q1 and Q2 in relation to the VR performance metric for the Driven and Field tests. Values closer to 1 or -1 respectively mean a strong direct or inverse relation between both metrics to correlate, while those close to 0 mean that there is no relation between them. We can see, in general, there was a positive correlation in all the cases. In relation to the Driven test, as expected, both correlations ($r_{Q1,\ VR}$ and $r_{Q2,\ VR}$) in NY and SYD were very strong, while correlations in BCN were the weakest. The reason for this lower correlation in the BCN case was that, on one hand, it had the worst VR values due to the high number of 3D buildings to be loaded, but, on the other hand, people maintained high interest and satisfaction when flying above Barcelona because it was the best known

Table 8
Results of question Q1 for the Field test.

Data Density	Joy	Desire	Fascin.	Satisfac.	Sadness	Disgust	Boredom	Disatis.
High	11.1%	11.1%	55.5%	22.2%	0%	0%	0%	0%
Medium	37.5%	12.5%	12.5%	60%	0%	0%	0%	0%
Low	22.2%	11.1%	11.1%	55.5%	0%	0%	0%	0%

Table 10
VR ranges for the different environments.

	VR ranges			
Data				
Density	0–5%	5–10%	10–15%	15–30%
High	66.7%	33.3%	0%	0%
Medium	75%	0%	25%	0%
Low	22.2%	22.2%	0%	55.5%

Table 11
Correlations for Q1, Q2 and VR.

Tests		Nr.Nodes	$r_{Q1,\ VR}$	$r_{Q2,\ VR}$
Driven	NY	3	0.65	0.62
		8	0.79	0.68
	BCN	3	0.15	0.30
		8	0.56	0.57
	SYD	3	0.52	0.63
		8	0.79	0.73
Field	High	5	0.63	0.74
	Medium	5	0.61	0.88
	Low	5	0.82	0.88

for the Spanish users. Likewise, the comparison of the two Driven tests with different node composition reveals that the correlation is stronger with 8 nodes for all the cases, especially for the $r_{Q1,\ VR}$ case. This is due to the fact that users had better immersive experience and as a consequence, they had more positive emotions (question Q1) and spent more time looking at the images. Regarding the Field test, we can see that a much better correlation was obtained given that in this case, the users navigated wherever they wanted and, as a consequence, all the answers achieved higher satisfaction values. Note that the highest correlation (near 90%) was obtained for the relation between Q2 and the VR parameter in the cases of low and medium density, which were the ones that had the lowest load timeout and, as a consequence, the user satisfaction with the load time was higher.

This strong correlation between user satisfaction (parameters Q1 and Q2) and the VR metric denotes that the modeling of the VR parameter for a specific LG configuration running Google Earth can give us a reference regarding the satisfaction level of a user in a given environment.

6. Modeling

Taking into account the above discussion and our goal of being able to estimate the suitability of a specific cluster display wall to reach certain user requirements, in this section, a theoretical schema used to model the behavior of the LG system running Google Earth is defined, so that we can estimate the theoretical VR value that could be obtained from the modeling of the cluster visualization wall and user activity.

In general, works reported for modeling performance in the HPC field are based on some application benchmarks running in specific computing architectures, such as the one reported in [26]. In this kind of works, the main goal is to predict, through the model, the estimated final execution time. However, in the present work, apart from the benchmark application (Google Earth) and the architecture (LG), the user perception is introduced as a key aspect to be taken into account when defining the final performance. In order to do this, we assume that the performance of the LG system is calculated according to the VR metric. So, there are two parts to be considered when calculating the VR: the time when the CPU is processing and the time when the CPU is idle. The time when the CPU is processing data, named Processing Time (T_{CPU}), is directly related to the infrastructure on which the application is running, together with the resolution of the image to be loaded. On the other hand, the time when the CPU is idle (T_{idle}) is more closely related to the user, who decides how much time he/she wants to spend visualizing the imagery. Note that the total visualization time (T_{total}) is the sum of T_{CPU} and T_{idle}, so it depends on the factors described above.

According to that reasoning, the model used to calculate the system performance is represented in Fig. 8, where two separated parts can be distinguished: the Machine and User modules. We can see that the inputs of these modules are separated into four key attributes:

- *Infrastructure*. We took two different aspects into account: a) the minimum power of the nodes in the cluster and the minimum broadband bandwidth, which are the most sensitive system parameters related to the LG performance [9], and b) the recommended minimum system requirements given by Google to run the Google Earth application [5]. Based on both points, we defined our reference cluster configuration in order to compare other systems to it. Our reference cluster was defined inside the following ranges: Number of nodes (from 3 up 8), CPU (from 1.6GHz to 3.0GHz), RAM (8GB at 1600MHz), Network Speed (from 6Mbps to 100Mbps) and Graphics Card (higher than NVIDIA GT620).
- *Data density*. The amount of data to be processed by the system is characterized by the number of calls to the *Draw function*. Note that this is given by the Google Earth application and is totally independent of the hardware.
- *Multimedia eXperience Level (MXL)*. When calculating the VR metric in previous tests, we concluded that the T_{idle} was different for every

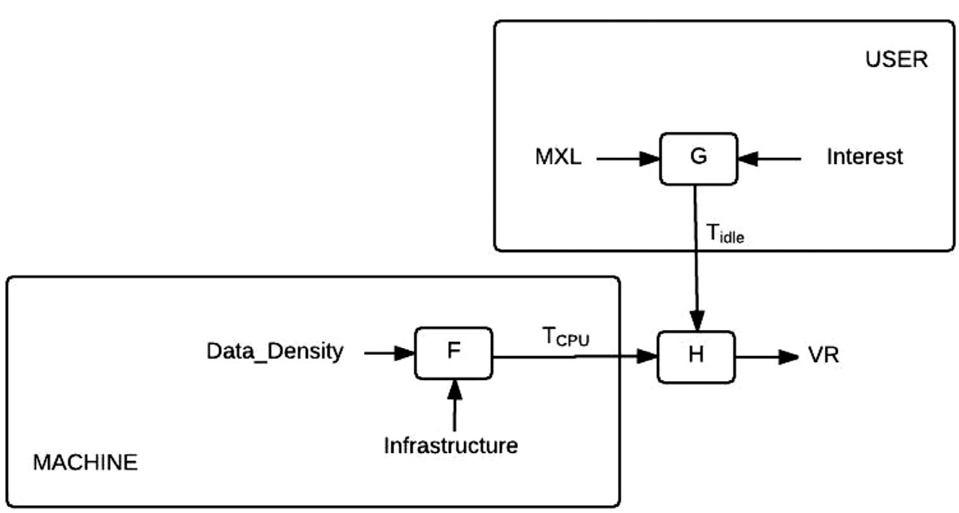


Fig. 8. Schema to model VR.

Table 12
Tours' metrics.

Tour	T_{CPU}		$Draw\_Calls$		T_{dl}	
Name	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
Sahara	120s	1.6s	154	0	7s	0.5s
Horsens	206s	2.3s	277	1	9s	0.6s
Alps	250s	1.4s	377	0	27s	0.3s
Barcelona	254s	3.1s	310	1	38s	1.4s
Venice	455s	3.9s	588	1	47s	2.4s
Paris	654s	4.7s	644	2	87s	3s

user who interacts with the system. While doing some tests, we noticed that the MXL or the ability of this kind of system was a significant parameter given that, when the MXL was higher, the T_{idle} was lower.

- *Interest.* The interest depends on how eager a user is to visualize a city or a place together with his MXL, which is reflected in the time that the system is idle. Taking previous experimentation into account, we consider "low interest" when the user has to complete an objective imposed by the facilitator of the test, whereas "high interest" includes the users who have freedom to navigate and are, therefore, more interested in the imagery.

6.1. Machine modeling

In this Section, we develop the modeling of the *Machine* module described in Fig. 8 by defining the T_{CPU} function as $F(Data\_density, Infrastructure)$. In order to do so, we carried out a test composed of six different tours with different size and polygon complexity. These are described in Table 12. The transition between each tour was done automatically using a very long timing jump to guarantee that the system was able to load the polygons and textures completely. All the tours were run in an LG cluster made up of eight nodes with a broadband connection of 100Mbps, which is inside our range for a reference cluster.

Table 12 shows the average and standard deviation obtained from running each tour 5 times focusing on the following parameters: T_{CPU} is the CPU time in seconds spent processing the images; $Draw\_calls$ is the number of Draw function calls needed to draw all the images and T_{dl} is the time needed to download the polygon structures and images, which is directly dependent on the size of these images. As can be observed, T_{CPU} average is much higher than T_{dl} in all the cases. Thus, we can assume that downloading an image and its processing is done simultaneously. So, if we take this assumption into account, we can state that the actual time used to call a single Draw function (T_{Draw}) can be defined as the rate between the time spent by the CPU to process all the multimedia (T_{CPU}) and the amount of data processed (number of Draw calls:

$$T_{Draw} = \frac{T_{CPU}}{Draw\_Calls} \tag{2}$$

To establish a reference value of T_{Draw}, we performed some tests with different configurations of the infrastructure. The tests were carried out using 24 configurations of clusters of eight nodes varying the CPU from 1.6GHz to 3GHz, and the broadband connection, from 6Mbps to 100Mbps, as shown in Table 13.

The time to compute a single Draw function (T_{Draw_ref}) for each configuration is shown in Table 13. These times show a clear difference for broadband connection below and above 40Mbps, which indicates that the broadband bandwidth is a bottleneck when its value is lower than this threshold. Therefore, the values of the results obtained for each system under the 40Mbps boundary were similar independently of the CPU power. On the other hand, the left side of the table shows the low influence of the connection bandwidth when the network resource

Table 13
Time per Draw function (T_{Draw_ref}).

	Broadband bandwidth					
Power of slowest node	100Mbps	80Mbps	40Mbps	20Mbps	10Mbps	6Mbps
3GHz	0.34s	0.36s	0.37s	0.71s	0.83s	1.03s
2.4GHz	0.39s	0.40s	0.42s	0.66s	0.84s	1.02s
2GHz	0.48s	0.50s	0.52s	0.64s	0.80s	1.02s
1.6GHz	0.52s	0.55s	0.56s	0.70s	0.83s	1.03s

becomes abundant. In this case, the CPU speed is the dominant parameter, where configurations with better CPUs achieved lower T_{Draw_ref} values.

It should be highlighted that these reference values were calculated keeping in mind the commodity hardware cluster architectures in the scope of the present paper. Considering a further step of using more powerful parallel systems with significantly higher CPU and communication capacities would imply the calculation of new appropriate reference values, while maintaining the proposed modeling process in the same way.

After having tested different configurations and obtained fixed values for every broadband connection, we can state that Eq. (3) can be used to calculate an approximation of the Processing Time (T_{CPU}).

$$T_{CPU} = Draw\_Calls \times T_{Draw\_ref} \tag{3}$$

where T_{Draw_ref} depends on the configuration of the broadband connection and is set by Table 13, while $Draw\_Calls$ depends on the imagery complexity and is given directly by Google Earth.

6.2. User modeling

This section explains how the user affects the T_{idle} by studying the user interest on the multimedia application and the skill or knowledge about using these applications (MXL).

In order to analyze the relationship between the MXL and the T_{idle} parameter, we took the experimentation carried out in Section 5 given that the users were classified by their MXL in those tests and this enables us to discriminate between them according to their skills.

Table 14 shows the average T_{idle} that the users achieved in those tests in relation to their MXL and interest. It is worth pointing out that we assume that users in the Driven test had Lower Interest (LI) than users from the Field test, who navigated wherever they wanted and showed Higher Interest (HI). It can be observed that the T_{idle} was higher when interest was also high.In the middle of the table the standard deviation for every MXL can be seen, denoted with the symbol σ. As can be observed, the values for almost every MXL can vary substantially. This is due to the fact that MXL is a very subjective perception and users were asked directly without any kind of objective evaluation regarding their technological skills. At the bottom of the table, the MXL factor for both high interest (MXL_{factor}_HI) and low interest (MXL_{factor}_LI) is shown. This was calculated as the relation between the T_{idle} of each MXLi and the MXL3 (used as a reference). The MXL_{factor} is used to weigh the influence of the MXL of each user in relation to the T_{idle}. According

Table 14
T_{idle} average according to MXL and interest.

	Interest	MXL1	MXL2	MXL3	MXL4	MXL5
T_{idle}	High Interest (HI)	59s	56s	32s	22s	17s
	Low Interest (LI)	36s	33s	13s	10s	6s
	$\sigma_{HighInterest}$	17s	10s	13s	9s	10s
	$\sigma_{LowInterest}$	2s	8s	12s	14s	9s
	MXL_{factor}_HI	1.84	1.75	1	0.69	0.53
	MXL_{factor}_LI	2.77	2.54	1	0.77	0.46

Table 15
T_{idle_ref} according to data density and interest for a reference cluster.

Interest	Data density	T_{idle_ref}
High	High	8s
High	Medium	10s
High	Low	19s
Low	High	8s
Low	Medium	8s
Low	Low	7s

to the MXL_{factor}, we can observe that the higher the interest, the lower the factor (with the exception of MXL5). Moreover, there is a clear trend that shows that users with a lower MXL had a higher T_{idle} and, thus, this increases the MXL_{factor} values.

Table 15 shows the T_{idle_ref} values, which correspond to the average of the T_{idle} for the Driven and Field tests achieved for any machine defined inside the range of our reference cluster. The values are categorized by the interest level (High (HI) or Low (LI)) of the user and the data density of the test. As with interest, the data density is categorized into the following ranges: High (above 160 MB), Medium (between 160 MB and 110 MB) or Low (below 110 MB) data density. The results show that, for High Interest, the average T_{idle_ref} significantly increases as the data density decreases from high to low. This is not the case for the Low Interest, as the value achieved is maintained at about 8s, which is due to the user trying to answer the questionnaire as soon as possible. Also, we can observe that the more interest the user has in the test, the longer he or she observes the imagery.

Thus, using the values from Table 15, we can define the T_{idle} for any set of polygons and textures with the following equation:

$$T_{idle} = T_{idle\_ref} \times MXL_{factor} \tag{4}$$

where T_{idle_ref} is the idle time for our reference machine and MXL_{factor} is defined using Table 14.

6.3. VR modeling

This Section combines the models of the *Machine* and *User* modules in order to obtain a mathematical model to calculate the value of the VR (VR_{Model}) in reference case scenarios.

Taking into account that VR is the relation between T_{idle} and T_{total} (see Eq. (1)), and that T_{total} is the sum of T_{CPU} and T_{idle}, Eq. (5) is used to calculate the VR_{Model}:

$$\begin{aligned} VR_{Model} &= \frac{T_{idle}}{T_{CPU} + T_{idle}} \\ &= \frac{T_{idle\_ref} \times MXL_{factor}}{Draw\_Calls \times T_{Draw\_ref} + T_{idle\_ref} \times MXL_{factor}} \end{aligned} \tag{5}$$

where T_{CPU} can be obtained using Eq. (3) by relating the *Draw_Calls*, given by Google Earth itself for a given imagery, and the T_{Draw} obtained from the values in Table 13. As for T_{idle}, the user interaction is taken into account and is calculated using Eq. (4) taking the T_{idle_ref} values from Table 15 and the MXL_{factor} from Table 14.

In order to compare our model with real values, Figs. 9a and b show the values of VR and VR_{Model} for both high MXL (3–5) and low MXL (1–2) respectively and from the VR average obtained in the two tests: Driven and Field tests. The axis of abscissas is classified in relation to the *Data Density - Interest* combination, where Data Density can be High, Medium or Low and Interest can be High or Low. As for Fig. 9a, it can be seen that the values of VR_{Model} approximately resemble the real VR obtained in the tests. By observing Fig. 9b, we can see that the values of VR_{Model} also have a trend similar to the real VR values, with the exception of the MH and LH points, but with more deviation. This behavior was observed previously in Section 5.1, where users with a low MXL achieved a VR with greater dispersion than the average VR. This is due to the categorization of the users' MXL, which is a subjective task and, thus a challenge that could be tackled in a new research line.

7. Conclusion

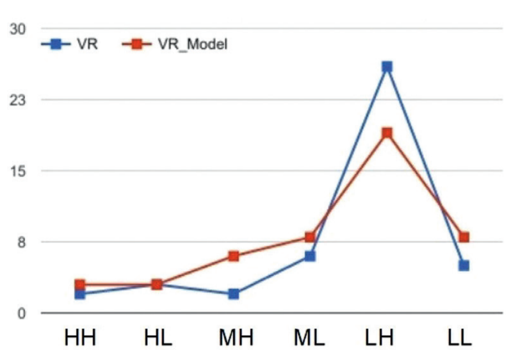
In this work, we carried out a performance study for a specific cluster display wall infrastructure named Liquid Galaxy and developed by Google, to obtain insight into how this performance affects user perception and enables a relationship to be established with some usability aspects (satisfaction, effectiveness and efficiency). The performance was measured with a metric, Visualization Rate (VR), that computes the CPU idle time in relation to the total CPU time. Google Earth was chosen as a representative case of an interactive application with high-performance requirements.

We studied the correlation between the VR and satisfaction of the users by performing two complementary usability tests with different node compositions of the Liquid Galaxy System:

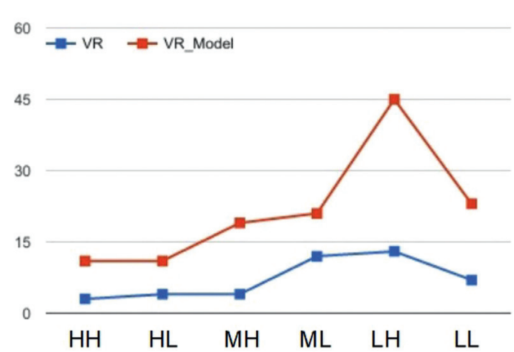
1. A Driven test, where participants followed a specific set of predefined tasks guided by a facilitator in a controlled space (similar to a usability lab), and
2. a Field test, in which customers of a travel agency navigated freely, following their own interests. The test was done in the agency's installations, mixing the experience with the activity of the employees and other clients, and so being a real interactive situation.

In all the cases, the post-task questionnaire answered by the users showed a direct and positive correlation of the VR metric with the usability parameters of effectiveness and efficiency. Additionally, it was shown that the correlation rose when the number of nodes was increased, due to the higher sensation of immersivity in the visualization process.

Although the users were positive influenced by the system itself



(a) VR for High MXL(3-5)



(b) VR for Low MXL(1-2)

Fig. 9. Values of VR and VR_{Model} for MXL3 to MXL5 (left) and MXL1 to MXL2 (right).

(due to its novelty), we found more satisfying answers when they were able to visit places of their own interest. Nevertheless, in both cases, a strong correlation was found between satisfaction usability parameter and VR. Thus, it enabled a relationship between performance and users' perception to be established.

Taking into account this relationship, we developed a mathematical model to calculate the expected VR value, called VR_{Model}, for a given cluster display wall infrastructure and users profile. The experimentation showed that the VR_{Model} was quite accurate and followed the same trend as the real VR obtained through the previous tests. Thus, this is a theoretical schema that is able to be extended according to the willingness of eventually evaluating different kinds of cluster visualization systems and user profiles.

References

[1] T. DeFanti, D. Acevedo, R. Ainsworth, M. Brown, S. Cutchin, G. Dawe, et al., The future of the CAVE, Cent. Eur. J. Eng. 1 (1) (2011) 16–37. Springer.
[2] A. Johson, J. Leigh, P. Morin, et al., Geowall: stereoscopic visualization for geoscience research and education, Comput. Graphics Appl. 26 (6) (2006) 10–14. IEEE.
[3] Nirnimesh, P. Harish, P. Naraayanan, Garuda: a scalable tiled display wall using commodity PCs, IEEE Trans. Visualization Comput. Graphics 13 (5) (2007) 864–877.
[4] Google, 2017, Liquid Galaxy Website: http://www.google.com/earth/explore/showcase/liquidgalaxy.html, retrieved at. 06–07.
[5] Google, 2017, Google Earth: http://www.google.com/earth/index.html, retrieved at. 06–07.
[6] K. Haynes, 2016, 4k Resolution: http://projects.ict.usc.edu/force/cominghome/cavesl/, retrieved at. 05–27.
[7] Openarena team, 2017a, OpenArena (Quake 3 Arena): http://www.openarena.ws, retrieved at. 01–07.
[8] WebGL apps, 2017b, WebGL Applications: https://webglsamples.org, retrieved at. 06–07.
[9] I. Arroyo, F. Giné, C. Roig, T. Granollers, Analyzing google earth application in a heterogeneous commodity cluster display wall, J. Multimedia Tools Appl. (2015) 1–26.
[10] ISO9241-210:2010, Ergonomics of Human System Interaction Part 210: Human-centered Design for Interactive Systems (formerly known as 13407).
[11] ISO/IEC 25010:2011, Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models.
[12] D.S. Tan, J.K. Stefanucci, D.R. Proffitt, et al., The infocockpit: providing location and place to aid human memory, in: Proceedings of the 2001 Workshop on Perceptive User Interfaces, ACM, 1–4. 10.1145/971478.971526.
[13] R. Ball, C. North, Effects of tiled high-resolution display on basic visualization and navigation tasks, In CHI'05 Extended Abstracts on Human Factors in Computing Systems, ACM, 2005, pp. 1196–1199.
[14] G. Humphreys, M. Eldridge, I. Buck, et al., WireGL: a scalable graphics system for clusters, Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, ACM, 2001, pp. 129–140.
[15] T. Ilmonen, M. Reunanen, P. Kontio, Broadcast GL: an alternative method for distributing openGL API calls to multiple rendering slaves, J. WSCG 13 (2) (2005) 65–72.
[16] B. Neal, P. Hunkin, A.M. Gregor, Distributed openGL rendering in network bandwidth constrained environments. Eurographics Symposium on Parallel Graphics and Visualization, (2011).
[17] B. Yost, Y. Haciahmetoglu, C. North, Beyond visual acuity: the perceptual scalability of information visualizations for large displays, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07), (2007), pp. 101–110, https://doi.org/10.1145/1240624.1240639.
[18] A. Endert, C. Andrews, Y.H. Lee, et al., Visual encodings that support physical navigation on large displays, Proceedings of Graphics Interface (GI '11), (2011), pp. 103–110.
[19] C. Andrews, A. Endert, C. North, Space to think: Large high resolution displays for sensemaking, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10), (2010), pp. 55–64, https://doi.org/10.1145/1753326.1753336.
[20] RA. Ruddle, R.G. Thomas, R.S. Randell, et al., Performance and interaction behaviour during visual search on large, high-resolution displays, Inf. Visualization 14 (2) (2015) 137–147, https://doi.org/10.1177/1473871613500978.
[21] K. Reda, A.E. Johnson, M.E. Papka, et al., Modeling and evaluating user behavior in exploratory visual analysis, Inf. Visualization 15 (4) (2016) 325–339, https://doi.org/10.1177/1473871616638546.
[22] ISO DIS, 9241-11 (1998): ergonomic requirements for office work with visual display terminals (VDTs)-part 11-guidance on usability, international organization for standardization, 1988.
[23] I. Arroyo, Cloud point visualization for liquid galaxy, 2016, https://summerofcode.withgoogle.com/projects/#6613839659401216, Google Summer of Code.
[24] J. Nielsen, Guerrilla HCI: using discount usability engineering to penetrate the intimidation barrier, Cost-justifying usability (1994) 245–272.
[25] Google., 2017, Tobii website: https://www.tobii.com/es/eye-tracking-research/global/products/hardware/tobii-glasses-eye-tracker, retrieved at. 09–16.
[26] C. Feichtinger, J. Habich, H. Kostler, et al., Performance modeling and analysis of heterogeneous lattice boltzmann simulations on CPU-GPU clusters, Parallel Comput. 46 (2015) 1–13, https://doi.org/10.1016/j.parco.2014.12.003.